Principal Component Analysis based on Robust Estimators of the Covariance or Correlation Matrix: Influence Functions and Efficiencies

Christophe Croux * Gentiane Haesbroeck †

Abstract

A robust principal component analysis can be easily performed by computing the eigenvalues and eigenvectors of a robust estimator of the covariance or correlation matrix. In this paper we derive the influence functions and the corresponding asymptotic variances for these robust estimators of eigenvalues and eigenvectors. The behavior of several of these estimators is investigated by a simulation study. Finally, the use of empirical influence functions is illustrated by a real data example.

Keywords: Influence Function, Principal Component Analysis, Robust Estimation, Robust Correlation Matrices.

AMS subject Classifications: 62H25, 62F35

*ECARE and Institut de Statistique, Université Libre de Bruxelles, CP-139, Av. F.D. Roosevelt 50, B-1050 Brussels, Belgium, ccroux@ulb.ac.be.

†FEGSS, University of Liège, Bd du Rectorat 7, B-4000 Liège, Belgium, G.Haesbroeck@ulg.ac.be

1 Introduction

One of the most popular techniques for analyzing multivariate data is principal component analysis (PCA). It consists of exploring the structure of a high-dimensional data set by projecting the observations onto the first principal components. These are obtained by computing the eigenvectors of the sample covariance or correlation matrix. The corresponding eigenvalues measure then the amount of information explained by the principal components. However, these estimators are extremely sensitive to outlying observations and conclusions drawn from contaminated principal components can be misleading. Several robustifications for PCA have been proposed (Jackson 1991, pages 365-371). Among these the replacement of the classical covariance or correlation matrix by a robust estimator is perhaps the most simple and intuitively appealing. Many simulation studies, starting with Devlin et al. (1981), have been carried out to find out which robust estimator should be used.

In this paper, a more formal comparison is undertaken by computing the influence functions for the estimators of the eigenvalues and eigenvectors. Corresponding asymptotic variances are also obtained. Results for M-estimators were already obtained by Jaupi and Saporta (1993), but our formulas are valid for any "regular" estimator, including high breakdown covariance matrix estimators. Several authors (Critchley 1985, Shi 1997) have suggested statistical diagnostics and graphical displays based on the influence function to detect influential points.

For every choice of the robust covariance matrix estimator, another robust PCA-method is obtained. An overview of existing estimators of multivariate location and scatter is given in Maronna and Yohai (1998). The words scatter matrix and covariance matrix will be abusively used as synonyms throughout this paper.

Three robust estimators (t_n, C_n) of multivariate location and scatter are considered in more detail: the M-estimator (Maronna 1976), the S-estimator (Rousseeuw and Leroy 1987, page 263, and Davies 1987) and the one-step reweighted Minimum Covariance Determinant (MCD) estimator (Rousseeuw 1985) which will be denoted by RMCD. Let us briefly review their definitions. Consider a sample of p-dimensional observations $x_1, \ldots, x_n$ and denote $d(x_i, t, C) = \sqrt{(x_i - t)^t C^{-1} (x_i - t)}$ the statistical distance between x_i and t, measured in the metric induced by the positive definite matrix C.

M-estimates are implicitly defined by

$$\begin{aligned} t_n &= \frac{\sum_{i=1}^n w_1(d(x_i,t_n,\mathrm{C}_n))x_i}{\sum_{i=1}^n w_1(d(x_i,t_n,\mathrm{C}_n))} \\ \mathrm{C}_n &= \frac{1}{n}\sum_{i=1}^n w_2(d^2(x_i,t_n,\mathrm{C}_n))(x_i-t_n)(x_i-t_n)^t \end{aligned}$$

where w_1 and w_2 are specified weight functions. Assuming monotonicity of w_2, Maronna (1976) showed that the M-estimation approach is less and less robust as the dimension increases since its breakdown point is at most $1/(p+1)$.

Unlike the M-estimators, the S-estimators belong to the class of high breakdown estimators of multivariate location and scatter. They are defined as the solutions (t_n, C_n) to the problem of minimizing $\det(C)$ subject to

$$\frac{1}{n}\sum_{i=1}^n \rho(d(x_i,t,C)) = b_0 \tag{1.1}$$

among all $(t,C) \in I\!R^p \times \mathrm{SPD}(p)$, with $\mathrm{SPD}(p)$ the set of all $p \times p$ symmetric and positive definite matrices. The constant b_0 is equal to $E_{F_0}\rho(\|x\|)$, with $F_0 = \mathrm{N}_p(0, \mathbf{I})$.

Finally, the one-step reweighted MCD estimators are defined as

$$\begin{aligned} t_n &= \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \\ \mathrm{C}_n &= c_1 \frac{\sum_{i=1}^n w_i (x_i - t_n)(x_i - t_n)^t}{\sum_{i=1}^n w_i} \end{aligned}$$

with $c_1 = (1-\delta)/F_{\chi^2_{p+2}}(q_\delta)$ a consistency factor and $q_\delta = \chi^2_{p,1-\delta}$ the upper δ-percent point of a χ^2_p distribution. The weights are computed as

$$w_i = \begin{cases} 1 & \text{if } \; d^2(x_i, t_n^0, \mathrm{C}_n^0) \le q_\delta \\ 0 & \text{otherwise} \end{cases}$$

where (t_n^0, C_n^0) are the initial MCD estimates. For defining this MCD estimator, consider all the subsets of size h $(\le n)$ from the sample and keep that subset whose covariance matrix has the smallest determinant. Then the location and scatter MCD estimates are given by the average and covariance matrix computed over this optimal subset. Typically, the size of the subset equals $h = [n(1-\alpha)]$, with $\alpha = 0.5$ or $\alpha = 0.25$. The breakdown points of MCD and RMCD are equal to $\alpha\%$.

When numerical values or graphical displays are given, they correspond to the following choice of functions and constants. For the M-estimator, the weight functions are chosen according to Huber's proposal:

$$w_1(y) = \frac{\psi_H(y, \sqrt{q_\tau})}{y} \quad \text{and} \quad w_2(y) = \frac{\psi_H(y, q_\tau)}{\beta\ y}.$$

where $\psi_H(y,k) = \max\{-k, \min\{y,k\}\}$ is Huber's psi function, β is a constant making the scatter estimate Fisher consistent at normal models and $q_\tau = \chi^2_{p,0.9}$. The function ρ in the definition of the S-estimator is the Biweight function $\rho(y) = \min(\frac{y^2}{2} - \frac{y^4}{2c_0^2} + \frac{y^6}{6c_0^4}, \frac{c_0^2}{6})$. To attain a breakdown point of 25%, c_0 is implicitly defined by $\rho(c_0) = \frac{b_0}{r}$ with $r = 0.25$. The breakdown point of the RMCD estimator will be the same as for the S-estimator, so $\alpha = 0.25$, and the trimming parameter δ equals 0.025 as suggested by Rousseeuw and Van Driessen (1997).

The outline of the paper is as follows. Section 2 presents influence functions and asymptotic variances of eigenvalues and eigenvectors computed from a robust covariance matrix while Section 3 presents similar results for the correlation matrix. In Section 4, some simulations are conducted to compare the performance of the robust estimators introduced above to estimate the PCA eigenvalues and eigenvectors. As a byproduct, a comparison of the estimated correlation coefficients is also reported. In Section 5, the use of the influence function as a data analytic tool is illustrated. Section 6 contains some conclusions.

2 Robust PCA based on the Covariance Matrix

Let $x_1, \ldots, x_n$ be an i.i.d. sample drawn from a p-variate distribution F. Throughout the paper F is assumed to be the normal distribution $\mathrm{N}(\mu, \Sigma)$ where $\mu \in I\!R^p$ and $\Sigma \in \mathrm{SPD}(p)$, the set of all $p \times p$ symmetric and positive definite matrices. Results can be easily shown to hold for any elliptically symmetric distribution F, but only for normal distributions one has that the principal components will be independent of each other (Hampel et al 1986, page 273). It is further supposed that Σ has distinct eigenvalues $\lambda_1 > \lambda_2 > \ldots > \lambda_p > 0$ with corresponding eigenvectors $v_1, v_2, \ldots, v_p$. The aim is to estimate these population eigenvalues and eigenvectors and to compute the corresponding influence functions. A generalization to multiple eigenvalues could be done as in Tanaka (1988).

An influence function is essentially the first derivative of the functional version of an estimator. Let $\mathcal{F}$ denote the set of all distributions on $\mathbb{R}^p$ (or a very large subset of it). A map $\mathrm{C} : \mathcal{F} \to \mathrm{SPD}(p)$ which sends an arbitrary distribution $G \in \mathcal{F}$ to $\mathrm{C}(G)$ is a statistical functional corresponding to an estimator C_n of Σ whenever $\mathrm{C}(F_n) = \mathrm{C}_n$, for every empirical distribution function F_n associated with observations $x_1, \ldots, x_n$. For example, the statistical functional defined as

$$\mathrm{C}(G) = E_G[(X - E_G[X])(X - E_G[X])^t]$$

for any distribution G having a second moment corresponds to the sample covariance matrix since

$$\mathrm{C}(F_n) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^t.$$

The notation $\mathrm{C}(X)$ instead of $\mathrm{C}(G)$ will be used whenever $X \sim G$. The functional representations of the eigenvectors and eigenvalues computed from C_n are denoted by $v_{\mathrm{C},j}$ and $\lambda_{\mathrm{C},j}$ for $j = 1, \ldots, p$. Of course, $v_{\mathrm{C},j}(G)$ and $\lambda_{\mathrm{C},j}(G)$ are just the eigenvectors and eigenvalues of $\mathrm{C}(G)$, for every $G \in \mathcal{F}$. At the empirical distribution function, $v_{\mathrm{C},j}(F_n) = v_{\mathrm{C}_n,j}$ and $\lambda_{\mathrm{C},j}(F_n) = \lambda_{\mathrm{C}_n,j}$. Throughout the paper, C is assumed to be Fisher consistent for Σ at F, i.e. $\mathrm{C}(F) = \Sigma$, and affine equivariant, meaning that

$$\mathrm{C}(AX + b) = A\mathrm{C}(X)A^t$$

for any $b \in \mathbb{R}^p$ and any $p \times p$ non singular matrix A. This implies immediately that Fisher consistency also holds for the eigenvector and eigenvalue functionals,

$$v_{\mathrm{C},j}(F) = v_j \text{ and } \lambda_{\mathrm{C},j}(F) = \lambda_j.$$

The functionals $v_{\mathrm{C},j}$ and $\lambda_{\mathrm{C},j}$ are orthogonal equivariant in the sense that

$$v_{\mathrm{C},j}(\Gamma X) = \Gamma v_{\mathrm{C},j}(X)$$

and

$$\lambda_{\mathrm{C},j}(\Gamma X) = \lambda_{\mathrm{C},j}(X)$$

for $j = 1, \ldots, p$ and for any $p \times p$ orthogonal matrix Γ.

To measure the robustness w.r.t. single outliers, it is common to compute their influence functions. By definition, the influence functions of $v_{\mathrm{C},j}$ and $\lambda_{\mathrm{C},j}$ are given by

$$\mathrm{IF}(x, v_{\mathrm{C},j}; F) = \lim_{\varepsilon \downarrow 0} \frac{v_{\mathrm{C},j}((1-\varepsilon)F + \varepsilon\Delta_x) - v_{\mathrm{C},j}(F)}{\varepsilon} \tag{2.1}$$

Table 1: Functions α_{C} for the M, S, RMCD and classical estimator of the covariance matrix.

M	$\alpha_{\mathrm{M}}(t) = \frac{p(p+2)w_2(t^2)}{p(p+2)+2E_{F_0}[w_2'(\|x\|^2)\|x\|^4]}$.
S	$\alpha_{\mathrm{S}}(t) = \frac{p\psi(t)}{\gamma_1 t}$ where $\psi(t) = \rho'(t)$ and $\gamma_1 = \frac{E_{F_0}[\psi'(\|y\|)\|y\|^2+(p+1)\psi(\|x\|)\|x\|]}{p+2}$.
RMCD	$\alpha_{\mathrm{RMCD}}(t) = \frac{d_2+2d_3}{d_2}\alpha_{\mathrm{MCD}}(t) + \frac{1}{d_2}I(t \leq \sqrt{q_\delta})$ where $q_\delta = \chi^2_{p,1-\delta}$, $d_2 = F_{\chi^2_{p+2}}(q_\delta)$, $d_3 = -\frac{1}{2}F_{\chi^2_{p+4}}(q_\delta)$ and $\alpha_{\mathrm{MCD}}(t) = \frac{I(t\leq\sqrt{q_\alpha})}{F_{\chi^2_{p+4}}(q_\alpha)}$ with $q_\alpha = \chi^2_{p,1-\alpha}$.
Cov	$\alpha_{\mathrm{Cov}}(t) = 1$

and

$$\mathrm{IF}(x, \lambda_{\mathrm{C},j}; F) = \lim_{\varepsilon \downarrow 0} \frac{\lambda_{\mathrm{C},j}((1-\varepsilon)F + \varepsilon\Delta_x) - \lambda_{\mathrm{C},j}(F)}{\varepsilon} \tag{2.2}$$

for $j = 1, \ldots, p$. The Dirac measure Δ_x is the distribution putting all its mass on x. For more details on influence functions and statistical functionals, see Hampel et al (1986). When the influence function of the scatter estimator is known, the influence functions (2.1) and (2.2) can be easily derived, as will be shown in Theorem 1 below. Before that, we give a lemma characterizing the general form of the influence function of a scatter matrix estimator. All the proofs are kept for the Appendix.

Lemma 1. *For any affine equivariant scatter matrix functional* C *possessing an influence function, there exist two functions* $\alpha_{\mathrm{C}}, \beta_{\mathrm{C}} : [0, \infty[\to I\!R$ *such that*

$$IF(x, \mathrm{C}; F) = \alpha_{\mathrm{C}}(d(x))(x-\mu)(x-\mu)^t - \beta_{\mathrm{C}}(d(x))\Sigma \tag{2.3}$$

with $d^2(x) = (x-\mu)^t\Sigma^{-1}(x-\mu)$ *and* $F = N_p(\mu, \Sigma)$.

From now on, only robust scatter matrix estimators possessing an influence function will be considered. Among them, focus is put on the estimators M, S and RMCD as defined in Section 1. The influence functions of these scatter estimators have been derived by Huber (1981, page 226) for the M-estimator and by Lopuhaä (1989 and 1997) for S and for reweighted estimators. The influence function of RMCD depends on the influence function of the initial MCD estimator which can be found in Croux and Haesbroeck (1998). The

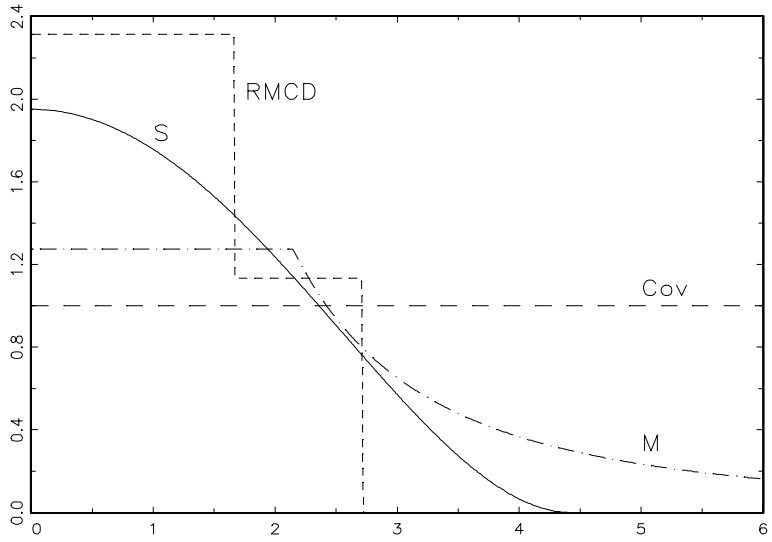


Figure 1. Examples of the function α_{C} for some robust estimators.

corresponding functions α_{C} are given in Table 1 and are plotted in Figure 1. The functions α_{S} and α_{M} are smooth while α_{RMCD} is a step-function with two discontinuities: one at $\sqrt{q_\alpha}$ which is due to the initial estimator and the other one at $\sqrt{q_\delta}$ resulting from the weighting scheme. Both $\alpha_{\mathrm{RMCD}}(t)$ and $\alpha_{\mathrm{S}}(t)$ become zero after a certain *rejection point* while $\alpha_{\mathrm{M}}(t)$ only redescends to zero at infinity. All these functions are non increasing meaning that their contribution to the influence function decreases as the distance between x and μ in the metric imposed by Σ increases. The function α_{Cov} is constant, implying that outliers are not given less weight.

Theorem 1. *Let F be a multivariate normal distribution with parameters μ and Σ. Define the scores of x as $z_k = v_k^t(x-\mu)$ for $k = 1, \ldots, p$ and let $d^2(x) = (x-\mu)^t\Sigma^{-1}(x-\mu)$. The influence functions of the eigenvectors and eigenvalues of* C *at F are then given by*

$$IF(x, \lambda_{\mathrm{C},j}; F) = \alpha_{\mathrm{C}}(d(x))z_j^2 - \beta_{\mathrm{C}}(d(x))\lambda_j$$

and

$$IF(x, v_{\mathrm{C},j}; F) = \alpha_{\mathrm{C}}(d(x)) \sum_{\substack{k=1 \\ k\neq j}}^{p} \frac{z_k z_j}{\lambda_j - \lambda_k} v_k$$

for $j = 1, \ldots, p$.

It follows now from Critchley (1985) that

$$\mathrm{IF}(x, v_{\mathrm{C},j}; F) = \alpha_{\mathrm{C}}(d(x))\mathrm{IF}(x, v_{\mathrm{Cov},j}; F). \tag{2.4}$$

The above equation confirms that the function α_{C} needs to be interpreted as a downweighting function. A redescending α_{C} function implies a bounded influence function for the eigenvectors.

In Figure 2, the influence function of the estimator $\lambda_{\mathrm{C},1}$ is plotted at $F = \mathrm{N}(0, \mathrm{diag}(2, 1))$. Figure 3 represents the norm of the influence function of the first eigenvector $v_{\mathrm{C},1}$. For C, both the classical covariance matrix and the S-estimator have been considered. The curves obtained for the S-estimator resemble the curves obtained for the classical estimator at the center of the distribution. Points further away are downweighted by the robust estimator while they can still have a large influence on the usual covariance matrix. For the eigenvalues, the most influential points are along the direction of the corresponding eigenvector. The norm of the influence function of the eigenvector is the largest along the bisectors.

The influence function can also be helpful for computing asymptotic variances. If a functional T corresponding to an estimator T_n is "sufficiently regular", then

$$\sqrt{n}\,(T_n - T(F)) \xrightarrow{d} \mathrm{N}_p(0, \mathrm{ASV}(T, F)) \tag{2.5}$$

with

$$\mathrm{ASV}(T, F) = E_F[\mathrm{IF}(x, T; F)\mathrm{IF}(x, T; F)^t] \tag{2.6}$$

(cfr. Hampel et al 1986, page 226). Taking (2.6) as definition of the asymptotic variance of a functional, the next corollary holds.

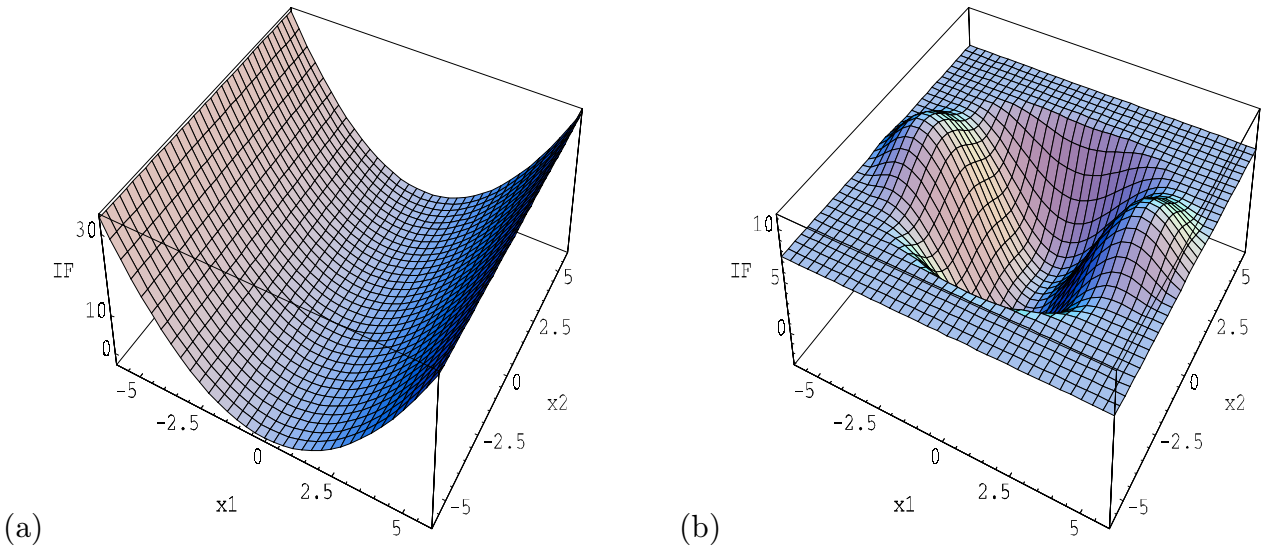
Corollary 1. *With the notations of Theorem 1, one has*

$$ASV(\lambda_{\mathrm{C},j}, F) = \lambda_j^2\, ASV(\mathrm{C}_{11}, F_0) \tag{2.7}$$

$$ASV(v_{\mathrm{C},j}, F) = ASV(\mathrm{C}_{12}, F_0) \sum_{\substack{k=1 \\ k \neq j}}^{p} \frac{\lambda_k \lambda_j}{(\lambda_j - \lambda_k)^2} v_k v_k^t, \tag{2.8}$$

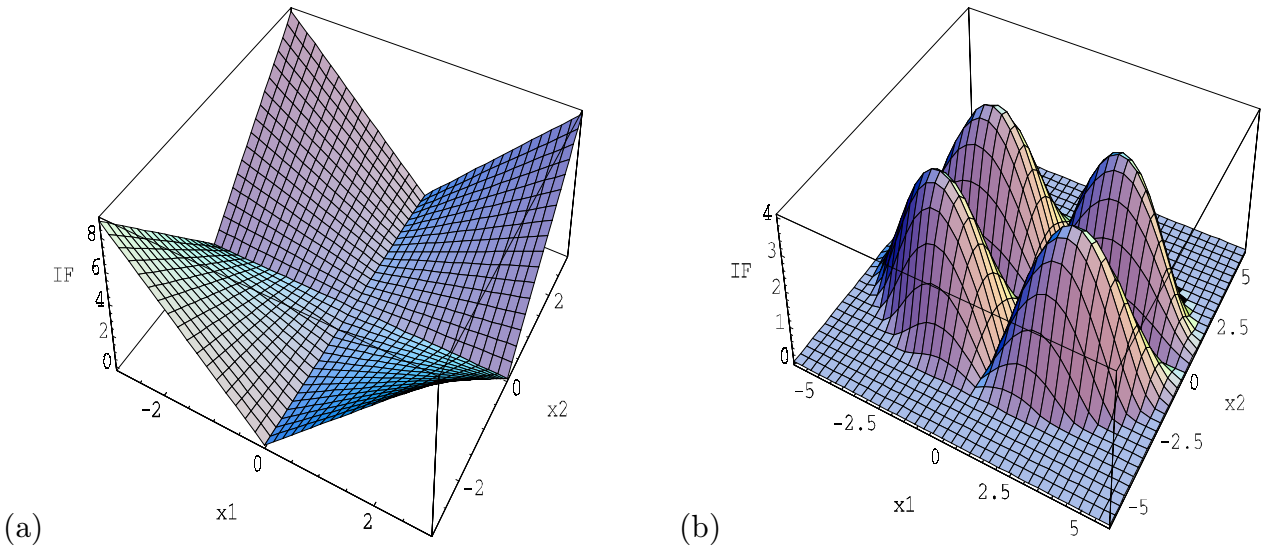
for $j = 1, \ldots, p$.

A formal proof of (2.5) for the robust eigenvectors and eigenvalues is beyond the scope of this paper. Boente (1987) provided regularity conditions in the case of M-estimators. Asymptotic



(a) (b)

Figure 2. Influence function of the largest eigenvalue for (a) the classical covariance matrix and (b) the S-estimator at $F = N_2(0, diag(2, 1))$.



(a) (b)

Figure 3. Norm of the influence function of the eigenvector corresponding to the largest eigenvalue for (a) the classical covariance matrix and (b) the S-estimator at $F = N_2(0, diag(2, 1))$.

Table 2: Asymptotic efficiencies at the normal distribution for the eigenvalue and eigenvector estimators based on the RMCD, S and M scatter matrix estimators.

		$p=2$	$p=3$	$p=5$	$p=10$	$p=30$
$\text{Eff}(\lambda_{\text{C},j}, F_0)$	M	0.881	0.895	0.947	0.974	0.991
	S	0.899	0.941	0.968	0.990	0.997
	RMCD	0.599	0.680	0.753	0.836	0.901
$\text{Eff}(v_{\text{C},j}, F_0)$	M	0.920	0.947	0.969	0.986	0.996
	S	0.850	0.924	0.967	0.988	0.997
	RMCD	0.635	0.742	0.820	0.873	0.933

theory for eigenvectors and eigenvalues of the sample covariance matrix is given by Anderson (1963). Asymptotic efficiencies at normal distributions can be defined as

$$\text{Eff}(\lambda_{\text{C},j}, F) = \frac{\text{ASV}(\lambda_{\text{Cov},j}, F)}{\text{ASV}(\lambda_{\text{C},j}, F)} = \frac{2}{\text{ASV}(\text{C}_{11}, F_0)}$$

and

$$\text{Eff}(v_{\text{C},j}, F) = \left(\frac{\det(\text{ASV}(v_{\text{Cov},j}, F))}{\det(\text{ASV}(v_{\text{C},j}, F))} \right)^{\frac{1}{p}} = \frac{1}{\text{ASV}(\text{C}_{12}, F_0)},$$

since the maximum likelihood estimators of (λ_j, v_j) at the normal model are given by the classical estimators (Joliffe 1986, page 41).

It is interesting to note that the efficiency of the eigenvalue estimators only depends on the efficiency of the diagonal elements of the scatter matrix estimator while the efficiency of the eigenvector estimators is computed from the efficiency of the off-diagonal elements. In Table 2, these efficiencies are reported for several values of p and for the estimators of interest. The reweighted MCD estimator results in the lowest efficiency values, even if they are not too bad. The S and M-estimators are comparable w.r.t. their efficiency and should be preferred to the classical covariance matrix since the small loss of efficiency is compensated by a better robustness.

The asymptotic variances (2.7) and (2.8) are often used to construct large sample confidence intervals. For example, a large sample $100(1-\alpha)$ % confidence interval for λ_j is

provided by

$$[\ \frac{\lambda_{\mathrm{C}_n,j}}{1+z_{\frac{\alpha}{2}}\sqrt{\frac{\mathrm{ASV}(\mathrm{C}_{11},F_0)}{n-1}}},\frac{\lambda_{\mathrm{C}_n,j}}{1-z_{\frac{\alpha}{2}}\sqrt{\frac{\mathrm{ASV}(\mathrm{C}_{11},F_0)}{n-1}}}\ ]$$

where $z_{\frac{\alpha}{2}}$ is the upper $100(\frac{\alpha}{2})$th percentile of a standard normal distribution. For the covariance matrix, $\mathrm{ASV}(\mathrm{Cov}_{11},F_0)=2$ and the usual formula is obtained again (cfr. Joliffe 1986, page 42).

3 Robust PCA based on the Correlation Matrix

It is well known that the principal components are not independent of the scales in which the original variables are measured. It is therefore often recommended to derive the principal components from the correlation matrix. This Section deals with the influence functions of the eigenvalues and eigenvectors computed from that matrix.

For any matrix $B\in I\!R^{p\times p}$, let $\mathrm{diag}(B)$ denote the $p\times p$ diagonal matrix whose elements are the diagonal elements of B. The population version of the correlation matrix is defined as $P=\Sigma_D^{-\frac{1}{2}}\Sigma\ \Sigma_D^{-\frac{1}{2}}$ where $\Sigma_D=\mathrm{diag}(\Sigma)$. A natural estimator for P is given by $\mathrm{R}_n=(\mathrm{C}_n)_D^{-\frac{1}{2}}\mathrm{C}_n\ (\mathrm{C}_n)_D^{-\frac{1}{2}}$ with $(\mathrm{C}_n)_D=\mathrm{diag}(\mathrm{C}_n)$. The corresponding functional is defined as $\mathrm{R}(G)=\mathrm{C}_D(G)^{-\frac{1}{2}}\mathrm{C}(G)\ \mathrm{C}_D^{-\frac{1}{2}}$, with $\mathrm{C}_D(G)=\mathrm{diag}(\mathrm{C}(G))$, for any distribution $G\in\mathcal{F}$. In this Section, $\lambda_1>\lambda_2>\ldots>\lambda_p>0$ denote the eigenvalues computed from P with corresponding eigenvectors $v_1,v_2,\ldots,v_p$. The notations $\lambda_{\mathrm{R},j}$ and $v_{\mathrm{R},j}$ are obvious.

The following lemma proves that the influence function of the functional R can be easily worked out when the function α_{C} appearing in the influence function of C is known. This lemma is then exploited to derive the influence function of the PCA functionals.

Lemma 2. *For any $x\in I\!R^p$, denote $\tilde{x}=\Sigma_D^{-\frac{1}{2}}(x-\mu)$ its standardized version and $D_{\tilde{x}}=diag(\tilde{x}\tilde{x}^t)$. The influence function of the functional* R *can be written as*

$$IF(x,\mathrm{R};F)=\alpha_{\mathrm{C}}(d(x))\{\tilde{x}\tilde{x}^t-(\frac{D_{\tilde{x}}P+PD_{\tilde{x}}}{2})\}$$

where $d^2(x)=(x-\mu)^t\Sigma^{-1}(x-\mu)$, $F=N_p(\mu,\Sigma)$ and α_{C} is the real-valued function of (2.3).

The influence function of the correlation functional can be rewritten in the closer form

$$\mathrm{IF}(x,\mathrm{R};F)=\alpha_{\mathrm{C}}(d(x))\mathrm{IF}(x,\mathrm{Corr};F) \tag{3.1}$$

where $\mathrm{IF}(x, \mathrm{Corr}; F)$ is the influence function of the ordinary correlation matrix as derived in Devlin et al (1975). Only the function α_{C} which already determined the form of $\mathrm{IF}(x, \mathrm{C}; F)$ appears in $\mathrm{IF}(x, \mathrm{R}; F)$. It may be easier to interpret the influence function when it is given element-wise. For the correlation functional, only the off-diagonal elements are of interest since $\mathrm{IF}(x, \mathrm{R}_{ii}; F) = 0$, as it should be. The element (i, j) with $i \neq j$ is given by

$$\mathrm{IF}(x, \mathrm{R}_{ij}; F) = \alpha_{\mathrm{C}}(d(x))\{\tilde{x}_i\tilde{x}_j - P_{ij}(\frac{\tilde{x}_i^2 + \tilde{x}_j^2}{2})\}$$

with $\tilde{x}_i = (x_i - \mu_i)/\sqrt{\Sigma_{ii}}$ $(i = 1, \ldots, p)$. Only the components i and j of x influence $\mathrm{IF}(x, \mathrm{Corr}_{ij}; F)$, but the other components may influence $d(x)$. When a robust estimator C is used, an extreme outlier in component k, which is however not outlying in two other components i and j, may have zero influence on R_{ij}. The influence functions of the correlation coefficients are of interest in their own right but are used here to derive the following Theorem.

Theorem 2. *The influence function of the eigenvalues and eigenvectors of* R *at the model distribution* $F = N(\mu, \Sigma)$ *are given by*

$$IF(x, \lambda_{\mathrm{R},j}; F) = \alpha_{\mathrm{C}}(d(x))\{\tilde{z}_j^2 - \lambda_j v_j^t D_{\tilde{x}} v_j\} \tag{3.2}$$

$$IF(x, v_{\mathrm{R},j}; F) = \alpha_{\mathrm{C}}(d(x)) \sum_{\substack{k=1 \\ k\neq j}}^{p} \left(\tilde{z}_k\tilde{z}_j - \frac{\lambda_k + \lambda_j}{2} v_j^t D_{\tilde{x}} v_k \right) \frac{v_k}{\lambda_j - \lambda_k} \tag{3.3}$$

where $\tilde{x} = \Sigma_D^{-\frac{1}{2}}(x - \mu)$, $d^2(x) = (x - \mu)^t\Sigma^{-1}(x - \mu)$ *and* $\tilde{z}_j$ *is the* j^{th} *coordinate of the standardized* x *in the basis of the eigenvectors, i.e.* $\tilde{z}_j = v_j^t\tilde{x}$.

Once again, the function α_{C} is responsible for the truncation of the influence functions derived for the usual correlation matrix since

$$\mathrm{IF}(x, \lambda_{\mathrm{R},j}; F) = \alpha_{\mathrm{C}}(d(x))\mathrm{IF}(x, \lambda_{\mathrm{Corr},j}; F)$$

and

$$\mathrm{IF}(x, v_{\mathrm{R},j}; F) = \alpha_{\mathrm{C}}(d(x))\mathrm{IF}(x, v_{\mathrm{Corr},j}; F).$$

Two minor remarks complete this Section. Firstly, it follows from $\mathrm{trace}(\mathrm{R}) = p$ that $\sum_{j=1}^{p} \mathrm{IF}(x, \lambda_{\mathrm{R},j}; F) = 0$. Secondly, the influence functions of the eigenvectors vanish for dimension $p = 2$.

4 Finite-sample Experiment

This Section uses simulations to compare the finite-sample performances of some robust estimators for estimating a correlation matrix and its principal components. The simulation set-up described in Devlin et al (1981) will be followed. The simulation consists of $m = 1000$ replications of 6-dimensional samples of $n = 50$ observations generated from four different distributions. The robust estimators involved in this study are the 25% breakdown point one-step reweighted MCD estimator, the 25% breakdown Biweight S-estimator as well as the Huber M-estimator defined in Section 1. Note that the M-estimator was already included in Devlin et al's paper.

The population correlation matrix will be $P = \begin{pmatrix} P_1 & 0 \\ 0 & P_2 \end{pmatrix}$, with $P_1 = \begin{pmatrix} 1 & & \\ 0.95 & 1 & \\ 0.30 & 0.10 & 1 \end{pmatrix}$ and $P_2 = \begin{pmatrix} 1 & & \\ -0.499 & 1 & \\ -0.499 & -0.499 & 1 \end{pmatrix}$.

The values of interest are the elements of the correlation matrix P and its eigenvalues ($\lambda_1 = 2.029, \lambda_2 = \lambda_3 = 1.499, \lambda_4 = 0.943, \lambda_5 = 0.028, \lambda_6 = 0.002$) and eigenvectors. The sampling distributions are taken as

1. The Normal distribution (NOR): $\mathrm{N}(0, P)$
2. A Symmetric Contaminated Normal (SCN) distribution: the mixture $0.9\ \mathrm{N}(0, P) + 0.1\ \mathrm{N}(0, 9P)$.
3. The multivariate Cauchy (CAU) which is defined as the distribution of $X = (\sqrt{S})^{-1}Y$, where $Y \sim \mathrm{N}(0, P)$ is independent of $S \sim \chi^2_1$.
4. An Asymmetric Contaminated Normal (ACN) distribution: the mixture $0.9\ \mathrm{N}(0, P) + 0.1\ \mathrm{N}(\mu, P)$, with (1) $\mu = 0.537 \times v_6$ or (2) $\mu = 50 \times v_6$ where v_6 is the eigenvector of P corresponding to λ_6. The case ACN(1) corresponds to intermediate outliers while ACN(2) generates extreme outliers. It will appear that the ACN(1) contamination is not heavy enough to let the classical estimator break down. This is the reason why the ACN(2) configuration, not considered in the study of Devlin et al, was introduced.

For computing the MCD estimator, the FAST-MCD algorithm of Rousseeuw and Van Driessen (1997) was applied, while the S-estimator was based on the SURREAL algorithm of Ruppert (1992). The iterative procedure given in Devlin et al (1981) for computing the M-estimator was used. All algorithms were implemented in GAUSS.

To assess the performance of the estimators of the elements of P, the finite-sample bias was computed as well as the Mean Squared Error. As in Devlin et al's study, the reported MSE's are defined as

$$\mathrm{MSE}(\rho_{ij}) = \frac{1}{m}\sum_{k=1}^{m}(\hat{\rho}_{ij}^{(k)} - \rho_{ij})^2$$

where $\rho_{ij} = \frac{1}{2}\ln\left(\frac{1+P_{ij}}{1-P_{ij}}\right)$ is the Fisher's z transform of the correlation coefficient P_{ij} and $\hat{\rho}_{ij}^{(k)}$ its estimate computed from the kth generated sample.

The biases of the different estimators are not reported since they were comparable across all sampling distributions (except for the M-estimator at the ACN-distributions, where the biases for the smaller correlations were higher). At the uncontaminated normal the classical correlation estimator is of course the most efficient, but the loss for the M and S-estimators is almost negligible. It can be seen that the Huber M-estimator outperforms the classical correlation at most other schemes. This explains why Devlin et al recommended the Huber M-estimator. At the ACN(2) configuration, one sees however that the M-estimator breaks down, confirming its lower breakdown point. This is not the case for the S-estimator, which appears to be the most robust at the asymmetric contamination distributions. Moreover, also for the other sampling schemes the S-estimator yields MSEs comparable to the M-estimator. The only exception is the Cauchy distribution, where the M-estimator behaves extremely well. The other competitor, RMCD, is clearly less efficient than the S-estimator, even at the contaminated distributions. As a first conclusion, one may say that the S-estimator seems to be the best estimator for the correlation coefficients.

The precision of the estimators for the eigenvalues of P was measured by

$$\mathrm{MSE}(\ln\lambda_i) = \frac{1}{m}\sum_{k=1}^{m}(\ln\hat{\lambda}_i^{(k)} - \ln\lambda_i)^2,$$

where $\hat{\lambda}_i^{(k)}$ is the estimate for the ith eigenvalue computed from the kth generated sample. Not very surprisingly, the same observations as for the correlation coefficients can be made from Table 3. The S-estimator turns out to be preferable, thanks to its relatively high

Table 3: $1,000\times$MSE of the $z-$transforms of the estimators of the elements of the correlation matrix and $100\times$MSE of the estimators of the logs of the eigenvalues under five different sampling schemes. The classical estimator is indicated by Corr, the Huber M-estimator by M, and the results based on the reweighted MCD and the S-estimator are in the two other columns.

dist	$1,000\times$ R_{ij}	$1,000\times$ MSE($\tilde{\rho}_{ij}$)				λ_i	$100\times$ MSE($\ln\lambda_i$)			
		Corr	M	S	RMCD		Corr	M	S	RMCD
	950	21	21	23	39	2.029	1	1	1	3
N	300	21	22	24	39	1.499	1	1	1	2
O	100	20	22	24	38	1.499	3	3	3	5
R	-499	22	23	21	40	0.943	3	3	4	12
	-499	24	25	23	43	0.028	8	9	10	20
	-499	21	22	22	39	0.002	8	8	8	22
	950	54	24	28	38	2.029	4	1	2	2
S	300	54	24	28	38	1.499	2	1	1	1
C	100	55	24	28	37	1.499	8	3	4	5
N	-499	56	26	27	37	0.943	13	4	5	10
	-499	55	27	29	36	0.028	23	10	13	17
	-499	57	25	27	35	0.002	22	9	11	19
	950	1128	34	63	71	2.029	43	2	5	5
C	300	1428	38	71	72	1.499	130	1	2	2
A	100	1400	38	71	70	1.499	481	5	9	10
U	-499	1400	35	65	69	0.943	877	8	19	22
	-499	1314	35	61	68	0.028	764	14	30	33
	-499	1361	35	62	67	0.002	1042	15	31	37
	950	21	22	25	36	2.029	1	1	1	2
A	300	21	22	22	38	1.499	1	1	1	1
C	100	20	22	23	38	1.499	3	3	3	5
N	-499	23	23	22	36	0.943	3	4	4	10
(1)	-499	23	24	23	33	0.028	5	6	7	17
	-499	21	22	22	36	0.002	567	458	381	18
	950	21	23	25	35	2.029	18	15	1	2
A	300	21	23	25	41	1.499	7	8	1	1
C	100	20	22	25	41	1.499	28	27	3	5
N	-499	8366	6170	24	36	0.943	1188	863	4	10
(2)	-499	8331	6139	23	34	0.028	17	13	11	17
	-499	8345	6150	24	31	0.002	368	585	9	19

efficiency combined with good robustness properties. There is one case where the S-estimator breaks down while RMCD does not: λ_6 for ACN(2). Apparently, the discontinuous character of RMCD (cfr. Figure 1) may lead to more robust solutions in some cases, at the price of a loss of efficiency.

It is also of interest to compare the estimators w.r.t. their performance to estimate the eigenvectors of the correlation matrix P. The estimations should be close to the true vector, i.e. the vectors v_j and $v_{\mathrm{C}_n,j}$ should be collinear. To measure their closeness, the cosine of the angle $\hat{\theta}_j$ they form is used. (For $j = 2$ and 3, $\hat{\theta}_j$ should be taken as the angle between $v_{\mathrm{C}_n,j}$ and its projection onto the space spanned by v_2 and v_3 since $\lambda_2 = \lambda_3$.) Figure 4 gives the empirical cumulative distribution functions of the realizations of $|\cos\hat{\theta}_j|$ over the $m = 1000$ replications under the distributions NOR and ACN for the M and S-estimators. Similar figures were given in Devlin et al (1981), but only at the NOR distribution. The distributions of $|\cos\hat{\theta}_5|$, $|\cos\hat{\theta}_3|$ and $|\cos\hat{\theta}_2|$ are omitted since the first one behaves like $|\cos\hat{\theta}_6|$, and the two others like $|\cos\hat{\theta}_4|$.

Comparing the two columns of Figure 4, one sees that the S-estimator is more robust than the M-estimator. Indeed, since the values of $|\cos\hat{\theta}_j|$ should be close to one, the cumulative distribution function should be peaked towards one. This is no longer true for the M-estimator at the ACN(2) distribution, while the S-estimator still finds good estimates for the eigenvectors in that case. The same exercise has been done for the RMCD estimator, yielding results which are even slightly better than for the S-estimator. Therefore, the RMCD estimator can still play its role in an exploratory analysis, when efficiency and inference issues are less important.

5 The Empirical Influence Function

Several authors (Critchley 1985, Shi 1997) have proposed local influence measures to detect influential points in a principal component analysis. Their measures were based, however, on the non-robust sample covariance matrix. Jaupi and Saporta (1993) introduced empirical influence measures based on M-estimators. In this Section the same approach is followed, now using a high breakdown estimator of multivariate scatter. If t_n and C_n are estimates obtained from the sample $x_1, \ldots, x_n$, then the empirical influence functions (EIF) for the

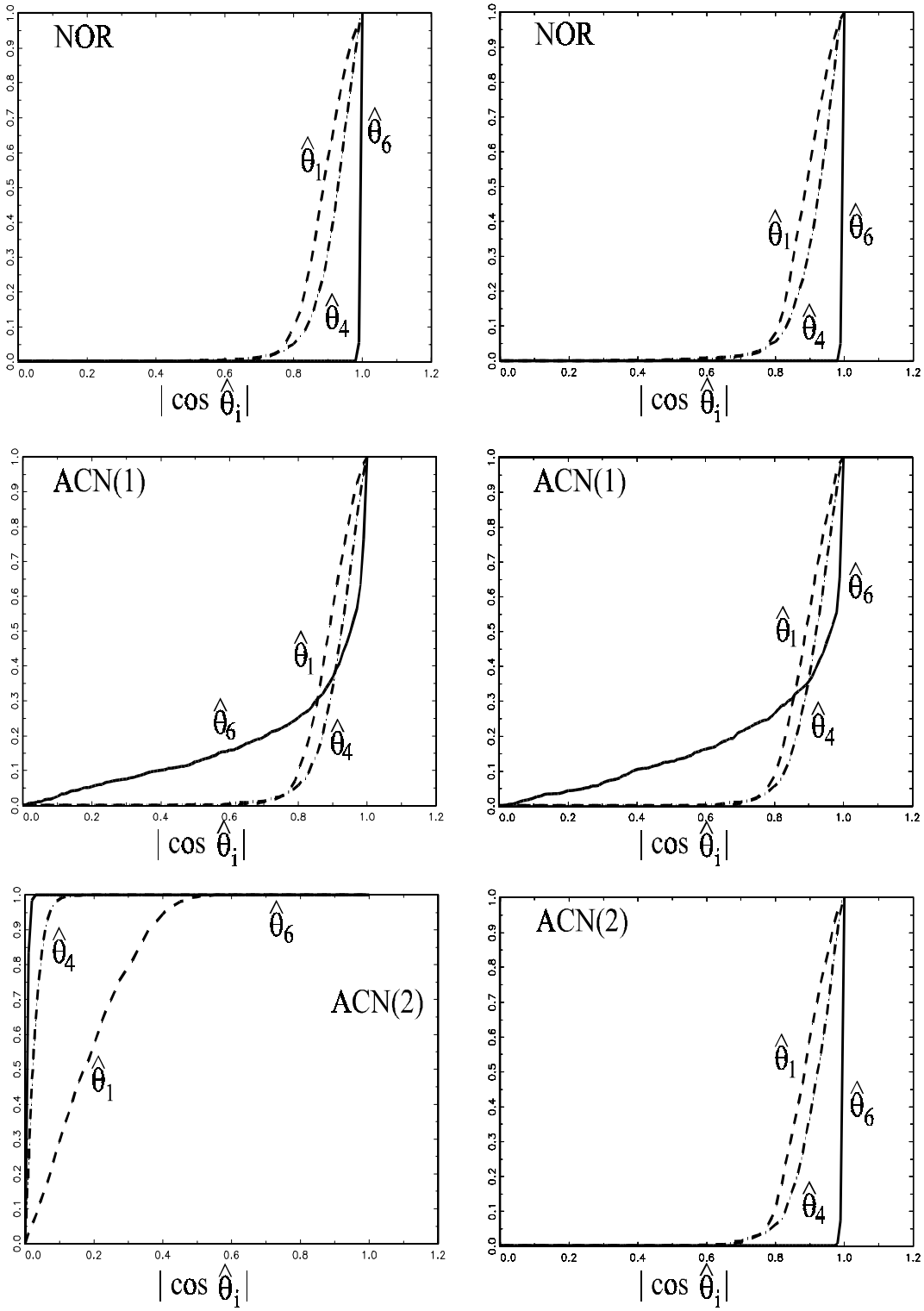


Figure 4. Simulated cumulative distribution functions for $|\cos\hat{\theta}_j|$, where $\hat{\theta}_j$ is the angle between the estimated and the population eigenvectors, for the M-estimator (first column) and the S-estimator (second column) under three different sampling schemes.

eigenvalues and eigenvectors are defined as

$$\mathrm{EIF}(x, \lambda_{\mathrm{C},j}) = \mathrm{IF}(x, \lambda_{\mathrm{C},j}; \hat{F}),$$

and

$$\mathrm{EIF}(x, v_{\mathrm{C},j}) = \mathrm{IF}(x, v_{\mathrm{C},j}; \hat{F})$$

for $j = 1, \ldots, p$ and $\hat{F} = \mathrm{N}(t_n, \mathrm{C}_n)$. Similar formulas apply for the correlation case.

Diagnostics which measure the influence of the observation x_i on the final estimates are then given by $\mathrm{EIF}(x_i, \lambda_{\mathrm{C},j})$ and $\mathrm{EIF}(x_i, v_{\mathrm{C},j})$. They may be visualized by plotting their values, or the norm of their values against the index of the observations. This is illustrated using the soil composition data set (20 observations on 4 variables) introduced by Kendall (1975), and used by all the papers mentioned before in this Section. In Figure 5, the value of $\mathrm{EIF}(x_i, \lambda_{\mathrm{R},j})$ is plotted for each observation with respect to its index, and this for every eigenvalue of the correlation matrix. The $\mathrm{EIF}(x_i, \lambda_{\mathrm{R},j})$ are computed once for the sample correlation matrix (left column), and once for the correlation estimator based on the RMCD (right column). The latter estimator was chosen because emphasis is more on exploring the data, then on inference. Moreover, the EIFs computed from S and M-estimators did not result in such a clear cut difference between the classical and robust approach.

Informal visual inspection of these plots show that observation 14 has a large influence on the second and third eigenvalues, while also observation 13 is quite influential on the first and fourth eigenvalues computed from the sample correlation matrix. This confirms the results obtained by the influence measures of Shi (1997). On the other hand, one notices that no observation has an influence which is much bigger than all the others on the eigenvalues of the RMCD correlation estimator. This is consistent with the philosophy of robustness, saying that a single observation may not influence too heavily the final estimate. Observations 13 and 14 have been downweighted by the RMCD-estimators, reason why their influence is greatly reduced.

To investigate the robustness of the index plots, contamination was introduced in the data set by changing the first and last component of the third observation (as was done in Jaupi and Saporta 1993). The "contaminated" index plots are represented on the same figures (dashed lines). First of all, notice that the EIF hardly changes for the robust RMCD estimator. The results for the classical estimator change quite a lot: one sees for example

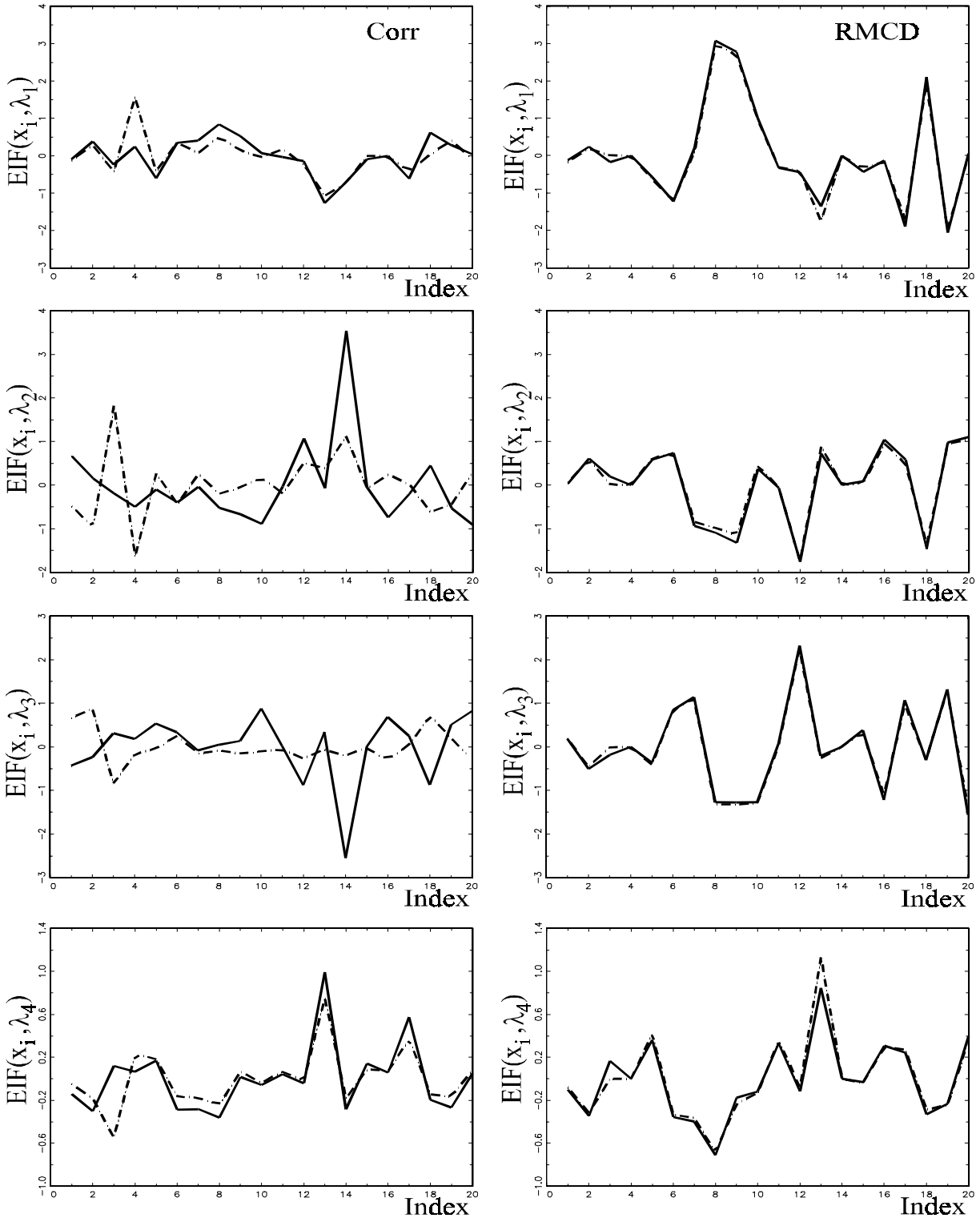


Figure 5. Empirical Influence Functions for the classical correlation matrix (left column) and for the RMCD estimator (right column) computed at the "soil"-data (solid line) and at the contaminated "soil"-data (dashed line).

that observation 14 is not an influential point anymore. It is also interesting to notice that the contaminated observation 3 does not appear to be extremely influential on the classical estimates. Of course, 3 is neither influential on the robust estimates since it has been downweighted to zero. A conclusion is that, although EIFs can be useful to determine which points are more influential that others, they are not suitable for outlier detection.

To detect outliers, it seems to be better to use the *Robust Distances*, which are robustified versions of the Mahalanobis distances

$$\mathrm{RD}_i = d(x_i, t_n, \mathrm{C}_n) = \sqrt{(x_i - t_n)^t \mathrm{C}_n^{-1} (x_i - t_n)}, \tag{5.1}$$

with (t_n, C_n) robust estimates of location and scatter. Observations with RD_i bigger than the critical value $\sqrt{\chi^2_{4,0.975}} = 3.34$ can be considered as potential outliers (cfr. Rousseeuw and van Zomeren, 1990). Mahalanobis distances (obtained by using the sample mean and covariance in (5.1)) and Robust Distances based on the RMCD estimator are reported for the most interesting observations in the table below:

	Soil Data					Contaminated Soil Data				
Index	3	4	12	13	14	3	4	12	13	14
Mahalanobis Distance	1.21	3.13	1.96	2.43	3.04	4.19	2.80	1.88	2.35	2.86
Robust Distance	1.19	4.52	2.75	2.36	4.25	29.9	4.45	2.67	2.31	4.12

Observation 4 has the largest distance for the clean data set, but is not detected as extremely influential for the eigenvalues. It was however detected by Shi (1997) as influential for the eigenvectors and it also comes up in the plot for λ_2 on the contaminated data. Notice that the classical Mahalanobis distance is not well suited for detecting outliers, since it is based on non robust estimators. Another tool to detect outliers is to make side by side boxplots of the scores on the robustly estimated principal components, as was suggested in (Croux and Ruiz, 1996).

6 Conclusion

Outlier resistant principal component estimators can be obtained by computing eigenvalues and eigenvectors of a robust estimate of the covariance or correlation matrix. Applications using M-estimators of scatter can be found in (Campbell 1980, Rivest and Plante 1988,

Daigle and Rivest 1992). Since the breakdown point of an M-estimator decreases towards zero with the dimension, high breakdown methods seem to be preferable. The most available high breakdown scatter matrix estimators are the *Minimum Volume Ellipsoid* (MVE) and the *Minimum Covariance Determinant* estimator (Rousseeuw 1985). Simulations have been conducted to study the behavior of principal components based on the latter estimators by Naga and Antille (1990), and Todorov, Neykov and Neytchev (1992). Although they can be very useful in a first stage of a data analysis, they lack statistical efficiency. (The MVE-estimator even has a non-normal convergence, reason why it was not included in the present study.) The theoretical results and simulations in this paper favor the use of S-estimators, since they combine high efficiency with appealing robustness properties, including a smooth influence function. In an exploratory data analysis, the RMCD-approach is a valuable alternative.

General expressions for influence functions have been given, which can be used for any scatter matrix estimator possessing an influence function. From them, asymptotic variances can be computed. It was also shown how influence functions can be used as an empirical diagnostic tool.

In case that the number of variables is bigger than $n(1-\alpha)$, the S and RMCD method with breakpoint α % are no longer applicable. (It can be seen that in this case, definition (1.1) and the definition of MCD yield an infinity of possible solutions). This is a serious drawback, since in many applications they are more variables than observations (e.g. Locantore et al (1999)). Projection based methods (Li and Chen 1985, Croux and Ruiz-Gazen 1996) may yield an outcome here.

Covariance and correlation matrices play a crucial role in many multivariate statistical techniques. Robustification of these techniques can be obtained using robust estimators for Σ and P. As such, Pison et al (1999) propose a robust way to factor analysis.

7 Appendix

Proof of Lemma 1: For $X \sim F$, denote F_0 the distribution of $\Sigma^{-\frac{1}{2}}(X-\mu)$. Since C is an affine equivariant functional,

$$\mathrm{C}(X) = \Sigma^{\frac{1}{2}}\mathrm{C}(\Sigma^{-\frac{1}{2}}(X-\mu))\Sigma^{\frac{1}{2}}$$

implying

$$\mathrm{IF}(x, \mathrm{C}; F) = \Sigma^{\frac{1}{2}} \mathrm{IF}(\Sigma^{-\frac{1}{2}}(x-\mu), \mathrm{C}; F_0)\Sigma^{\frac{1}{2}}. \tag{7.1}$$

Since F_0 is spherically symmetric, Lemma 1, page 276 in Hampel et al (1986) guarantees that two real-valued functions α_{C} and β_{C} exist such that

$$\mathrm{IF}(u, \mathrm{C}; F_0) = \alpha_{\mathrm{C}}(\|u\|)uu^t - \beta_{\mathrm{C}}(\|u\|)\mathbf{I}_p \tag{7.2}$$

where $\mathbf{I}_p$ is the $p \times p$ identity matrix and $\|.\|$ indicates the Euclidean norm. Substituting (7.2) in (7.1) yields

$$\mathrm{IF}(x, \mathrm{C}; F) = \Sigma^{\frac{1}{2}} \left\{ \alpha_{\mathrm{C}}(\|\Sigma^{-\frac{1}{2}}(x-\mu)\|)\Sigma^{-\frac{1}{2}}(x-\mu)(x-\mu)^t\Sigma^{-\frac{1}{2}} - \beta_{\mathrm{C}}(\|\Sigma^{-\frac{1}{2}}(x-\mu)\|)\mathbf{I}_p \right\} \Sigma^{\frac{1}{2}}.$$

Equation (2.3) follows now immediately by noting that $\|\Sigma^{-\frac{1}{2}}(x-\mu)\| = \{(x-\mu)^t\Sigma^{-1}(x-\mu)\}^{\frac{1}{2}} = d(x)$. □

The proof of Theorem 1 relies on the following lemma, which mimics Lemma 2.1 of Sibson (1979) and is included for reasons of completeness:

Lemma 3. *Let $S : \mathcal{F} \to PDS(p)$ be a statistical functional and F a p-dimensional distribution. Suppose that $IF(x, S; F)$ exists and let $S(F) = \Xi$. Denote $v_1, \ldots, v_p$ and $\lambda_1, \ldots, \lambda_p$ the eigenvectors and eigenvalues of Ξ. Then the influence functions of $v_{S,j}$ and $\lambda_{S,j}$ $(j = 1, \ldots, p)$ are given by*

$$IF(x, \lambda_{S,j}; F) = v_j^t IF(x, S; F)v_j \tag{7.3}$$

and

$$IF(x, v_{S,j}; F) = \sum_{\substack{k=1 \\ k \neq j}}^{p} \frac{1}{\lambda_j - \lambda_k} \left(v_k^t IF(x, S; F)v_j \right) v_k. \tag{7.4}$$

Proof of Lemma 3: Since $\mathrm{S}(F_\varepsilon)v_{\mathrm{S},j}(F_\varepsilon) = \lambda_{\mathrm{S},j}(F_\varepsilon)v_{\mathrm{S},j}(F_\varepsilon)$ for $F_\varepsilon = (1-\varepsilon)F + \varepsilon\Delta_x$, simple derivation yields

$$\frac{\partial \mathrm{S}(F_\varepsilon)}{\partial \varepsilon}|_{\varepsilon=0}\ v_j + \Xi \frac{\partial v_{\mathrm{S},j}(F_\varepsilon)}{\partial \varepsilon}|_{\varepsilon=0} = \frac{\partial \lambda_{\mathrm{S},j}(F_\varepsilon)}{\partial \varepsilon}|_{\varepsilon=0}\ v_j + \lambda_j \frac{\partial v_{\mathrm{S},j}(F_\varepsilon)}{\partial \varepsilon}|_{\varepsilon=0}.$$

Rearranging the terms gives

$$(\Xi - \lambda_j \mathbf{I}_p)\mathrm{IF}(x, v_{\mathrm{S},j}; F) = (\mathrm{IF}(x, \lambda_{\mathrm{S},j}; F)\mathbf{I}_p - \mathrm{IF}(x, \mathrm{S}; F))v_j$$

which can, using $\Xi = \sum_{k=1}^{p} \lambda_k v_k v_k^t$ and $\mathbf{I}_p = \sum_{k=1}^{p} v_k v_k^t$, be written as

$$\sum_{\substack{k=1 \\ k\neq j}}^{p} [(\lambda_k - \lambda_j) v_k^t \mathrm{IF}(x, v_{\mathrm{S},j}; F)] v_k = -\sum_{\substack{k=1 \\ k\neq j}}^{p} \left[v_k^t \mathrm{IF}(x, \mathrm{S}; F) v_j\right] v_k + \left[\mathrm{IF}(x, \lambda_{\mathrm{S},j}; F) - v_j^t \mathrm{IF}(x, \mathrm{S}; F) v_j\right] v_j \tag{7.5}$$

Since $v_1, \ldots, v_p$ form an orthogonal basis, (7.5) implies (7.3) and

$$v_k^t \mathrm{IF}(x, v_{\mathrm{S},j}; F) = v_k^t \mathrm{IF}(x, \mathrm{S}; F) v_j / (\lambda_j - \lambda_k).$$

Noting that the side-condition $v_j^t v_j = 0$ implies that $\mathrm{IF}(x, v_{\mathrm{S},j}; F)$ has no component in the direction of v_j, (7.4) follows. □

Proof of Theorem 1: From Lemma 1, there exist two real-valued functions α_{C} and β_{C} such that

$$v_k^t \mathrm{IF}(x, \mathrm{C}; F) v_j = \alpha_{\mathrm{C}}(d(x))(v_k^t(x-\mu))((x-\mu)^t v_j) - \beta_{\mathrm{C}}(d(x)) v_k^t \Sigma v_j. \tag{7.6}$$

Lemma 3 combined with (7.6) and the equality $v_k^t \Sigma v_j = \lambda_j \delta_{jk}$ results in

$$\mathrm{IF}(x, \lambda_{\mathrm{C},j}; F) = v_j^t \mathrm{IF}(x, \mathrm{C}; F) v_j = \alpha_{\mathrm{C}}(d(x))(v_j^t(x-\mu))^2 - \beta_{\mathrm{C}}(d(x))\lambda_j$$

and

$$\mathrm{IF}(x, v_{\mathrm{C},j}; F) = \sum_{\substack{k=1 \\ k\neq j}}^{p} \frac{1}{\lambda_j - \lambda_k} \alpha_{\mathrm{C}}(d(x))(v_k^t(x-\mu))(v_j^t(x-\mu))^t v_k.$$

Replacing $v_j^t(x-\mu)$ by the z-score z_j yields the stated expressions. □

Proof of Corollary 1: Using (2.6) and Theorem 1, the asymptotic variance of $\lambda_{\mathrm{C},j}$ can be computed as:

$$\mathrm{ASV}(\lambda_{\mathrm{C},j}, F) = E_F[\mathrm{IF}(x, \lambda_{\mathrm{C},j}; F)^2] = E_F[\left(\alpha_{\mathrm{C}}(d(x)) z_j^2 - \beta_{\mathrm{C}}(d(x))\lambda_j\right)^2],$$

where $z_j = v_j^t(x-\mu)$.

With $u_j = z_j/\sqrt{\lambda_j}$, one has that $u = (u_1, \ldots, u_p)^t \sim F_0$. Moreover, $d^2(x) = \sum_{j=1}^{p} u_j^2$ and Lemma 1 imply

$$\begin{aligned} \mathrm{ASV}(\lambda_{\mathrm{C},j}, F) &= E_{F_0}[\left(\alpha_{\mathrm{C}}(\|u\|)\lambda_j u_j^2 - \beta_{\mathrm{C}}(\|u\|)\lambda_j\right)^2] \\ &= \lambda_j^2 E_{F_0}[\mathrm{IF}(u, \mathrm{C}_{jj}; F_0)^2] = \lambda_j^2 \,\mathrm{ASV}(\mathrm{C}_{11}, F_0) \end{aligned}$$

as $\mathrm{ASV}(\mathrm{C}_{jj}, F_0) = \mathrm{ASV}(\mathrm{C}_{11}; F_0)$ by symmetry. For the eigenvector estimator, the asymptotic variance is given by

$$\begin{aligned}\mathrm{ASV}(v_{\mathrm{C},j}, F) &= E_F[\mathrm{IF}(x, v_{\mathrm{C},j}; F)\mathrm{IF}(x, v_{\mathrm{C},j}; F)^t] \\ &= \sum_{\substack{k=1 \\ k\neq j}}^{p}\sum_{\substack{l=1 \\ l\neq j}}^{p} \frac{1}{\lambda_j - \lambda_k}\frac{1}{\lambda_j - \lambda_l} E_F[\alpha_{\mathrm{C}}(d(x))^2 z_k z_l z_j^2] v_k v_l^t. \end{aligned} \tag{7.7}$$

Using again the transformation $z_j/\sqrt{\lambda_j} = u_j$ to compute the expectation in (7.7) leads to

$$E_F[\alpha_{\mathrm{C}}(d(x))^2 z_k z_l z_j^2] = \sqrt{\lambda_k}\sqrt{\lambda_l}\lambda_j E_{F_0}[\alpha_{\mathrm{C}}(\|u\|)^2 u_k u_l u_j^2] = \lambda_k\lambda_j E_{F_0}[\alpha_{\mathrm{C}}(\|u\|)^2 u_k^2 u_j^2]\delta_{kl}. \tag{7.8}$$

Substituting (7.8) in (7.7) and using Lemma 1 yields

$$\begin{aligned}\mathrm{ASV}(v_{\mathrm{C},j}, F) &= \sum_{\substack{k=1 \\ k\neq j}}^{p} \frac{\lambda_j\lambda_k}{(\lambda_j - \lambda_k)^2} E_{F_0}[\mathrm{IF}(u, \mathrm{C}_{jk}; F_0)^2] v_k v_k^t \\ &= \mathrm{ASV}(\mathrm{C}_{12}, F_0) \sum_{\substack{k=1 \\ k\neq j}}^{p} \frac{\lambda_j\lambda_k}{(\lambda_j - \lambda_k)^2} v_k v_k^t \end{aligned}$$

since $\mathrm{ASV}(\mathrm{C}_{jk}, F_0) = \mathrm{ASV}(\mathrm{C}_{12}, F_0)$ by symmetry. □

Proof of Lemma 2: Since $\mathrm{R}(F_\varepsilon) = \mathrm{C}_D^{-\frac{1}{2}}(F_\varepsilon)\mathrm{C}(F_\varepsilon)\mathrm{C}_D^{-\frac{1}{2}}(F_\varepsilon)$ with $F_\varepsilon = (1-\varepsilon)F + \varepsilon\Delta_x$, $\mathrm{C}(F) = \Sigma$ and $\mathrm{C}_D(F) = \Sigma_D$, derivation yields

$$\mathrm{IF}(x, \mathrm{R}; F) = -\frac{1}{2}\{\mathrm{IF}(x, \mathrm{C}_D; F)\Sigma_D^{-1}R + R\Sigma_D^{-1}\mathrm{IF}(x, \mathrm{C}_D; F)\} + \Sigma_D^{-\frac{1}{2}}\mathrm{IF}(x, \mathrm{C}; F)\Sigma_D^{-\frac{1}{2}}. \tag{7.9}$$

¿From Lemma 1, there exist two functions α_{C} and β_{C} such that

$$\mathrm{IF}(x, \mathrm{C}_D; F) = \alpha_{\mathrm{C}}(d(x))D_x - \beta_{\mathrm{C}}(d(x))\Sigma_D \tag{7.10}$$

with $D_x = \mathrm{diag}((x-\mu)(x-\mu)^t)$. Inserting (2.3) and (7.10) in (7.9) yields

$$\mathrm{IF}(x, \mathrm{R}; F) = -\frac{1}{2}\alpha_{\mathrm{C}}(d(x))(D_x\Sigma_D^{-1}R + R\Sigma_D^{-1}D_x) + \alpha_{\mathrm{C}}(d(x))(\Sigma_D^{-\frac{1}{2}}(x-\mu))(\Sigma_D^{-\frac{1}{2}}(x-\mu))^t.$$

Putting $\tilde{x} = \Sigma_D^{-\frac{1}{2}}(x-\mu)$ and $D_x\Sigma_D^{-1} = \Sigma_D^{-1}D_x = D_{\tilde{x}}$ completes the proof. □

Proof of Theorem 2: From Lemma 2, it follows that

$$v_j^t\mathrm{IF}(x, \mathrm{R}; F)v_k = \alpha_{\mathrm{C}}(d(x))\left\{(v_j^t\tilde{x})(v_k^t\tilde{x})^t - \left(\frac{v_j^t D_{\tilde{x}} R v_k + (Rv_j)^t D_{\tilde{x}} v_k}{2}\right)\right\}.$$

With $Rv_k = \lambda_k v_k$ and $Rv_j = \lambda_j v_j$, Lemma 3 gives

$$\mathrm{IF}(x, \lambda_{\mathrm{R},j}; F) = v_j^t \mathrm{IF}(x, \mathrm{R}; F) v_j = \alpha_{\mathrm{C}}(d(x))(\tilde{z}_j^2 - \lambda_j v_j^t D_{\tilde{x}} v_j)$$

and

$$\mathrm{IF}(x, v_{\mathrm{R},j}; F) = \alpha_{\mathrm{C}}(d(x)) \sum_{\substack{k=1 \\ k \neq j}}^{p} \left(\tilde{z}_k \tilde{z}_j - \frac{\lambda_k + \lambda_j}{2} v_j^t D_{\tilde{x}} v_k \right) \frac{v_k}{\lambda_j - \lambda_k}.$$

□

8 References

Anderson, T.W. (1963), "Asymptotic Theory for Principal Components Analysis", *The Annals of Mathematical Statistics*, 34, 122–148.

Boente, G. (1987), "Asymptotic Theory for Robust principal Components", *Journal of Multivariate Analysis,* 21, 67-78.

Campbell, N.A. (1980), "Robust Procedures in Multivariate Analysis: Robust Covariance Estimation", *Applied Statistics*, 29, 5–14.

Critchley, F. (1985), "Influence in Principal Component Analysis", *Biometrika*, 72, 627–636.

Croux, C., and Haesbroeck, G. (1998), "Influence Function and Efficiency of the Minimum Covariance Determinant Scatter Matrix Estimator", to appear in *The Journal of Multivariate Analysis.*

Croux, C. and Ruiz-Gazen, A. (1996), "A Fast Algorithm for Robust Principal Components based on Projection Pursuit," in *Compstat: Proceedings in Computational Statistics*, ed.A. Prat, Heidelberg: Physica-Verlag. 211–217.

Daigle, G., and Rivest, L-P. (1992), "A Robust Biplot", *The Canadian Journal of Statistics,* 120, 241–255.

Davies, P.L. (1987), "Asymptotic Behavior of S-Estimators of Multivariate Location Parameters and Dispersion Matrices", *The Annals of Statistics,* 15, 1269–1292.

Devlin, S.J., Gnanadesikan, R., and Kettenring, J.R. (1975), "Robust Estimation and Outlier Detection with Correlation Coefficients", *Biometrika*, 62, 531–545.

Devlin, S.J., Gnanadesikan, R., and Kettenring, J.R. (1981), "Robust Estimation of Dispersion Matrices and Principal Components", *Journal of the American Statistical Association*, 76, 354–362.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel, W.A. (1996), *Robust Statistics: The Approach based on Influence Functions*, New York: Wiley.

Huber, P.J. (1981), *Robust Statistics*, New York: Wiley.

Jackson, J.E. (1991), *A User's Guide to Principal Components*, New York: Wiley.

Jaupi, L., and Saporta, G. (1993), "Using the Influence Function in Robust Principal Components Analysis", *New Directions in Statistical Data Analysis and Robustness*, eds. S. Morgenthaler, E. Ronchetti, and W.A. Stahel, Basel: Birkhäuser, 147–156.

Joliffe, I.T. (1986), *Principal Component Analysis*, New York: Springer-Verlag.

Kendall, M. (1975), *Multivariate Analysis*, Charles Griffin & Company LTD.

Li, G. and Chen, Z (1985), "Projection-Pursuit Approach to Robust Dispersion Matrices and Principal Components : Primary Theory and Monte Carlo", *Journal of the American Statistical Association*, 80, 759-766.

Locantore, N., Marron, J.S., Simpson, D.G., Tripoli, N., Zhang, J.T., and Cohen, K.L. (1999), "Robust Principal Components for Functional Data," to appear in *Test*.

Lopuhaä, H.P. (1989), "On the Relation Between S-estimators and M-estimators of Multivariate Location and Covariance", *The Annals of Statistics*, 17, 1662–1683.

Lopuhaä, H.P. (1997), "Asymptotics of Reweighted Estimators of Multivariate Location and Scatter", preprint TU Delft.

Maronna, R.A. (1976), "Robust M-estimators of Multivariate Location and Scatter", *The Annals of Statistics*, 4, 51–67.

Maronna, R.A., and Yohai (1998), "Robust Estimation of Multivariate Location and Scatter", in *Encyclopedia of Statistical Sciences Update Volume 2*, eds. S. Kotz, C. Read, and D. Banks, New York: Wiley, 589–596.

Naga, R., and Antille, G. (1990), "Stability of Robust and Non-Robust Principal Component Analysis," *Computational Statistics & Data Analysis,* 10, 169–174.

Pison, G., Rousseeuw, P.J., Filzmoser, P. and Croux, C. (1999), "Factor Analysis in a Robust Way", preprint University of Antwerp.

Rivest, L-P., and Plante, N. (1988), "L'Analyse en Composantes Principales Robuste", *Revue Statistique Appliquée*, XXXVI (1), 55–66.

Rousseeuw, P.J. (1985), "Multivariate Estimation with High Breakdown Point", in *Mathematical Statistics and Applications, Vol. B*, eds. W. Grossmann, G. Pflug, I. Vincze, and W. Wertz, Dordrecht: Reidel, 283–297.

Rousseeuw, P.J., and Leroy, A.M. (1987), *Robust Regression and Outlier Detection*, New York: Wiley.

Rousseeuw, P.J., and Van Driessen, K. (1997), "A Fast Algorithm for the Minimum Covariance Determinant Estimator", preprint University of Antwerp.

Rousseeuw, P.J., and van Zomeren, B.C. (1990), "Unmasking Multivariate Outliers and Leverage Points", *Journal of the American Statistical Association,* 85, 633–639.

Ruppert, D. (1992), "Computing S-Estimators for Regression and Multivariate Location/Dispersion", *Journal of Computational and Graphical Statistics*, 1, 253–270.

Shi, L. (1997), "Local Influence in Principal Components Analysis", *Biometrika*, 84, 175–186.

Sibson, R. (1979), "Studies in the Robustness of Multidimensional Scaling: Perturbational Analysis of Classical Scaling", *Journal of the Royal Statistical Society B*, 41, 217–229.

Tanaka, Y. (1988), "Sensitivity Analysis in PCA: Influence on the Subspace Spanned by Principal Components", *Communications in Statistics: Theory and Methods*, 17, 3157–3175.

Todorov, V.K., Neykov N., and Neytchev, P.N. (1992), "Stability of (High Breakdown Point) Robust Principal Components Analysis," in *COMPSTAT 1994, Short Communications in Computational Statistics,* eds. R. Dutter and W. Grossmann, 90–92.